

Static Search:

An Archivable and Sustainable Search Engine for the Digital Humanities

Joseph Takeda

(Digital Humanities Innovation Lab at Simon Fraser University)

Martin Holmes

(Humanities Computing and Media Centre at the University of Victoria)

[State the problem]

[Introduce the solution]

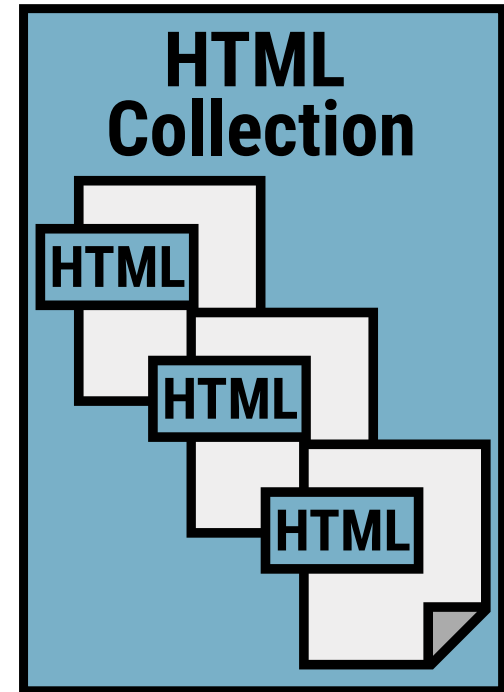
Demonstration applications

- <https://johnkeats.uvic.ca/search.html>
- <https://dvpp.uvic.ca/search.html>

How it works

You have:

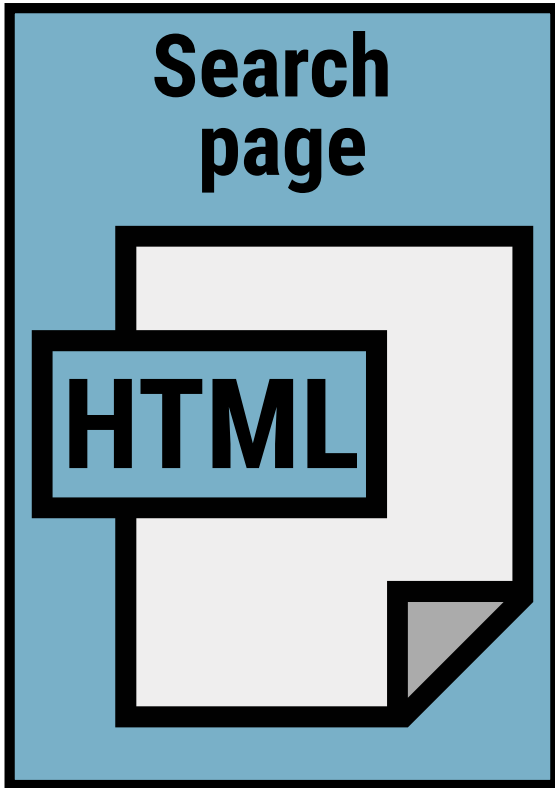
- A collection of HTML files
- All well-formed & valid XHTML



You add:

- An HTML search page containing a special div:

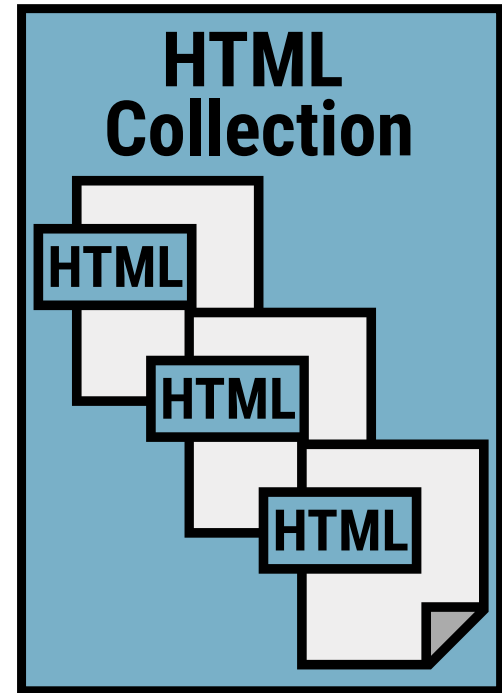
```
<div id="staticSearch"></div>
```



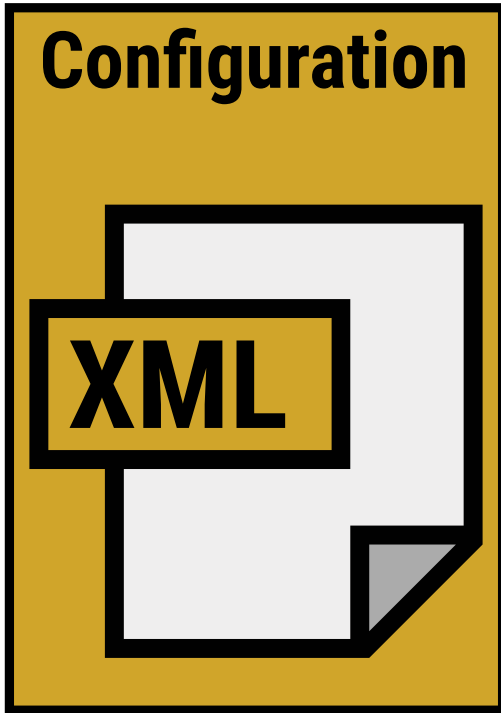
You add:

- Some metadata elements to support search filters:

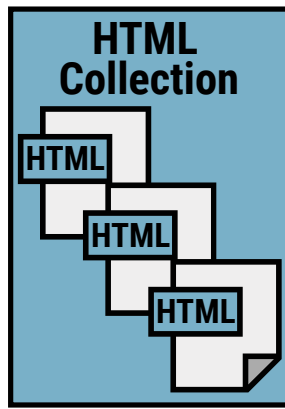
```
<meta name="Document type"  
class="staticSearch.desc"  
content="Poems" />
```

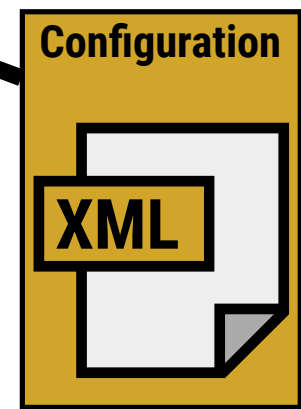
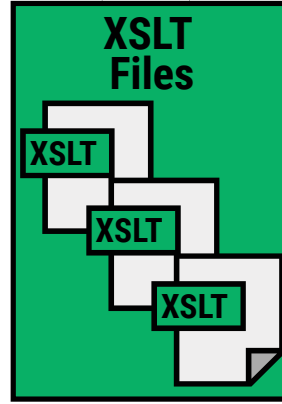
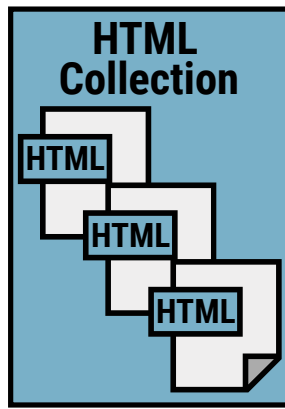


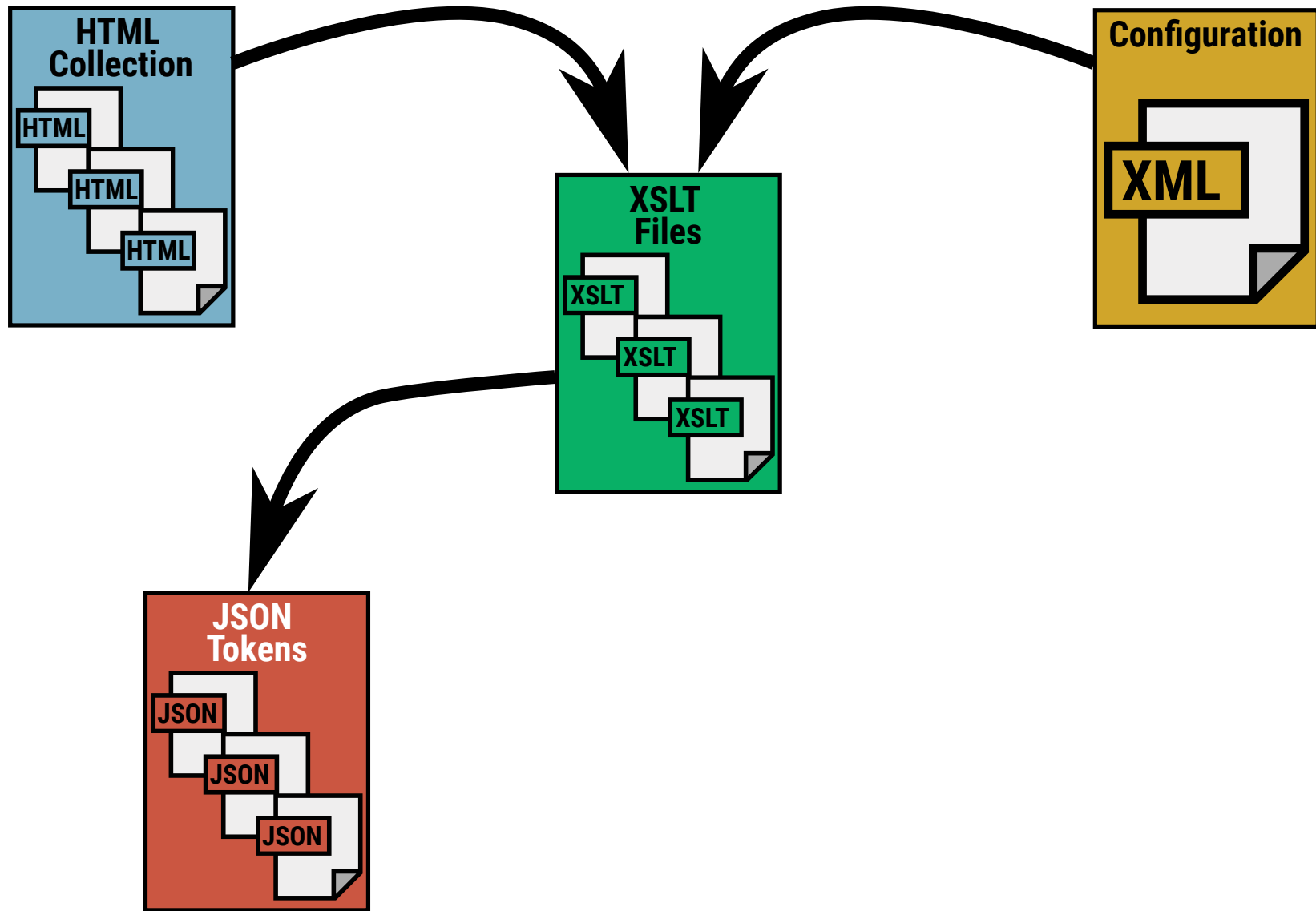
You create:

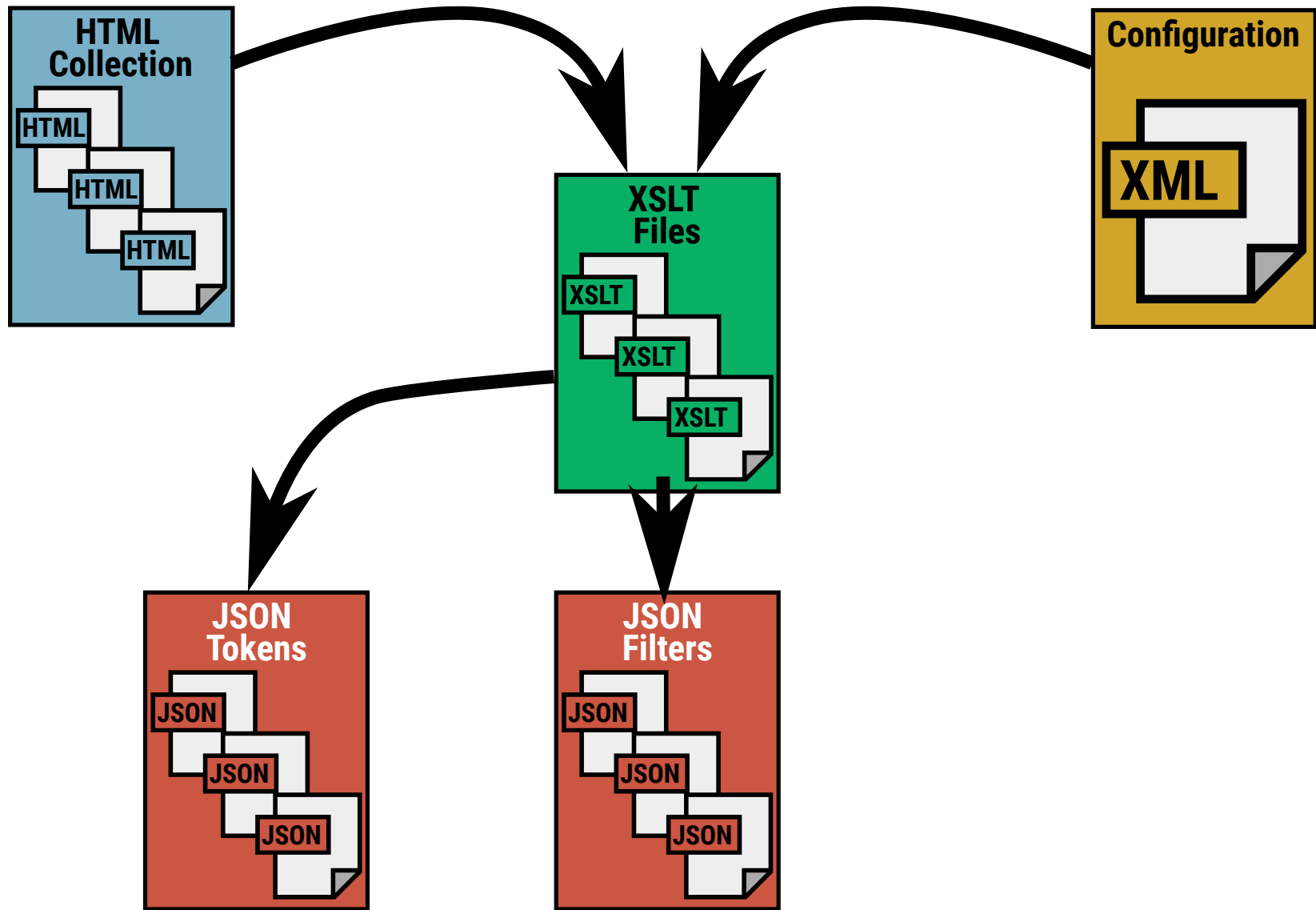


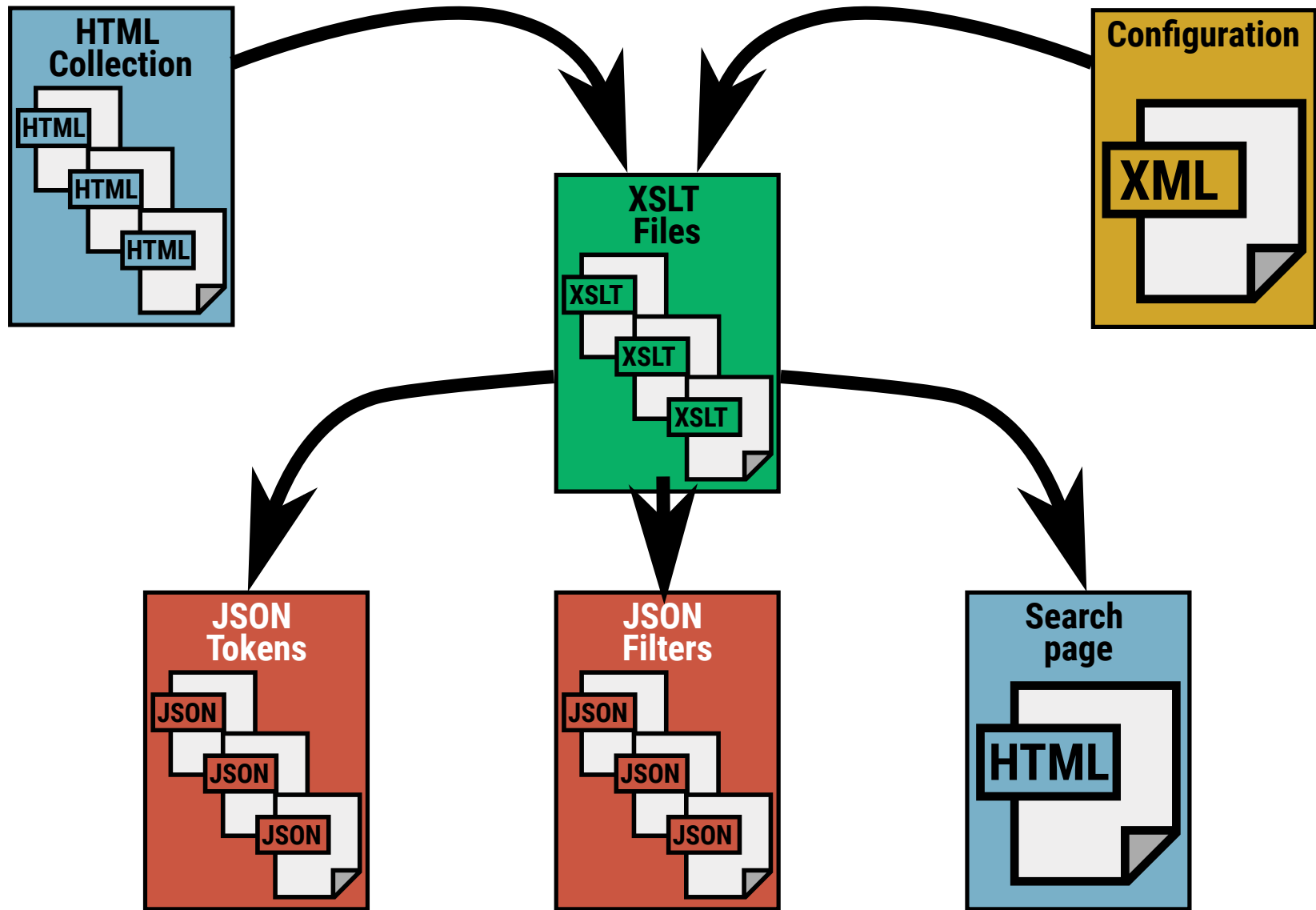
- A configuration file (XML) to specify the features and constraints for your search engine











JSON token file

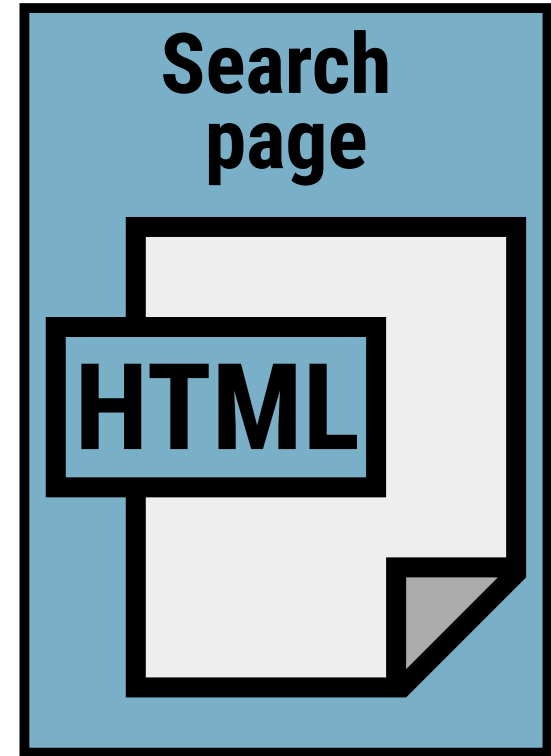
```
{
  "token": "unprofit",
  "instances": [
    {
      "docId": "pom_2025_ithe_old_man_of_hoy",
      "docUri": "poems/goodwords/1870/pom_2025_ithe_old_man_of_hoy.html",
      "score": 1,
      "contexts": [
        {
          "form": "unprofitably",
          "context": "...whole that day was spent <mark>unprofitably</mark>.",
          "weight": 1,
          "pos": 400
        }
      ]
    },
    {
      "docId": "pom_8733_john_and_joan_canto_ii",
      "docUri": "poems/blackwoods/1820/pom_8733_john_and_joan_canto_ii.html",
      "score": 1,
      "contexts": [
        {
          "form": "unprofitable",
          "context": "...too much ap- propriated unto <mark>unprofitable</mark> jocularities and facetiousness. Craving licence,...",
          "weight": 1,
          "pos": 164
        }
      ]
    }
  ]
}
```

Filter JSON

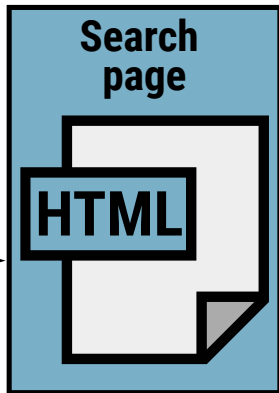
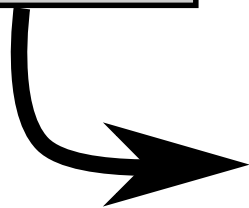
```
{
  "filterId": "ssBool2",
  "filterName": "Unsigned",
  "ssBool2_1": {
    "value": "true",
    "docs": [
      "poems/chambers_series/1867/pom_7594_the_husbands_request.html",
      "poems/alltheyearround/1875/pom_4271_the_hourglass.html",
      "poems/alltheyearround/1879/pom_4507_in_the_conservatory.html",
      "poems/blackwoods/1843/pom_9963_jolly_father_joe.html",
      "poems/blackwoods/1829/pom_10314_the_watchmans_lament.html"
    ]
  }
}
```

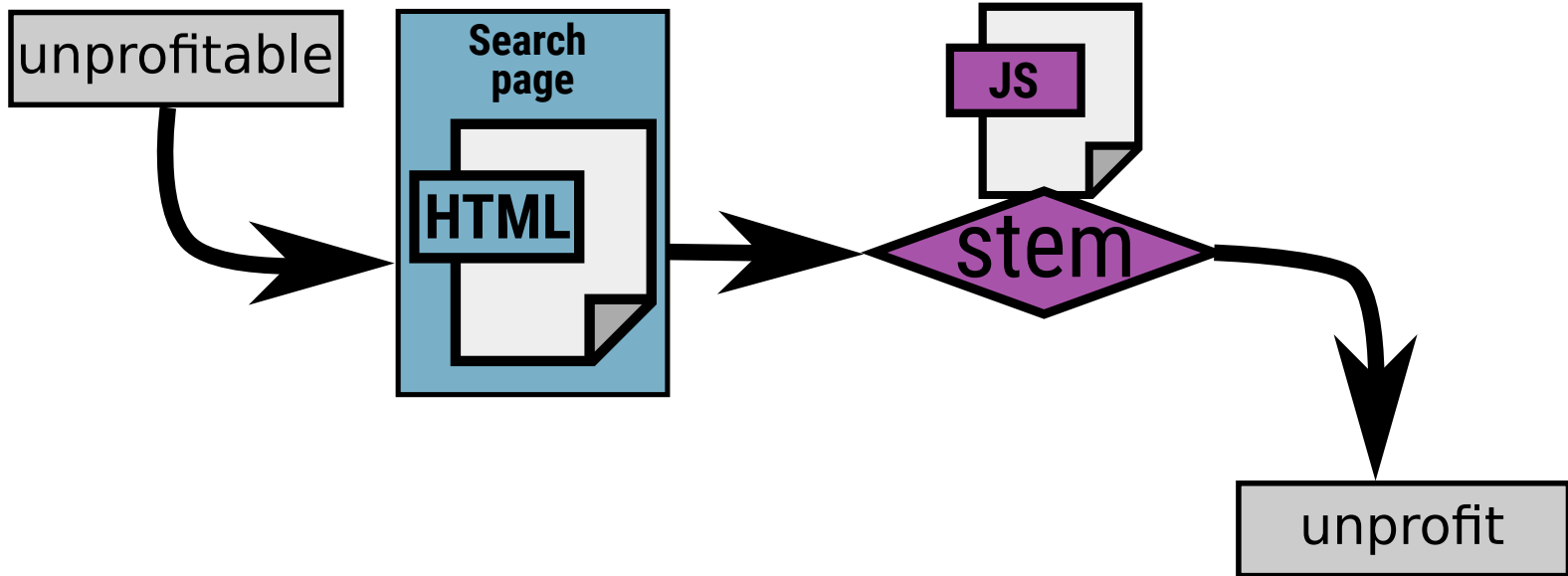

On the website...

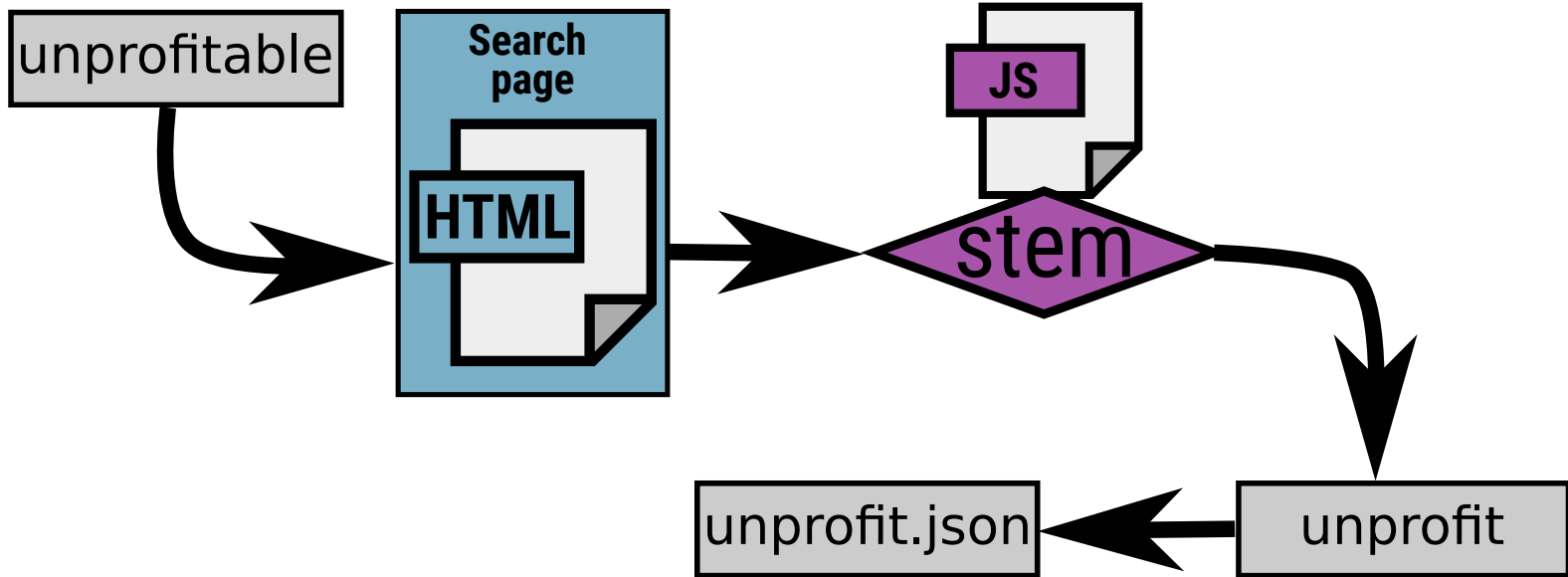
- ...of the thousands of JSON files...
- ...the search page JS retrieves only the ones it needs for the search you do:

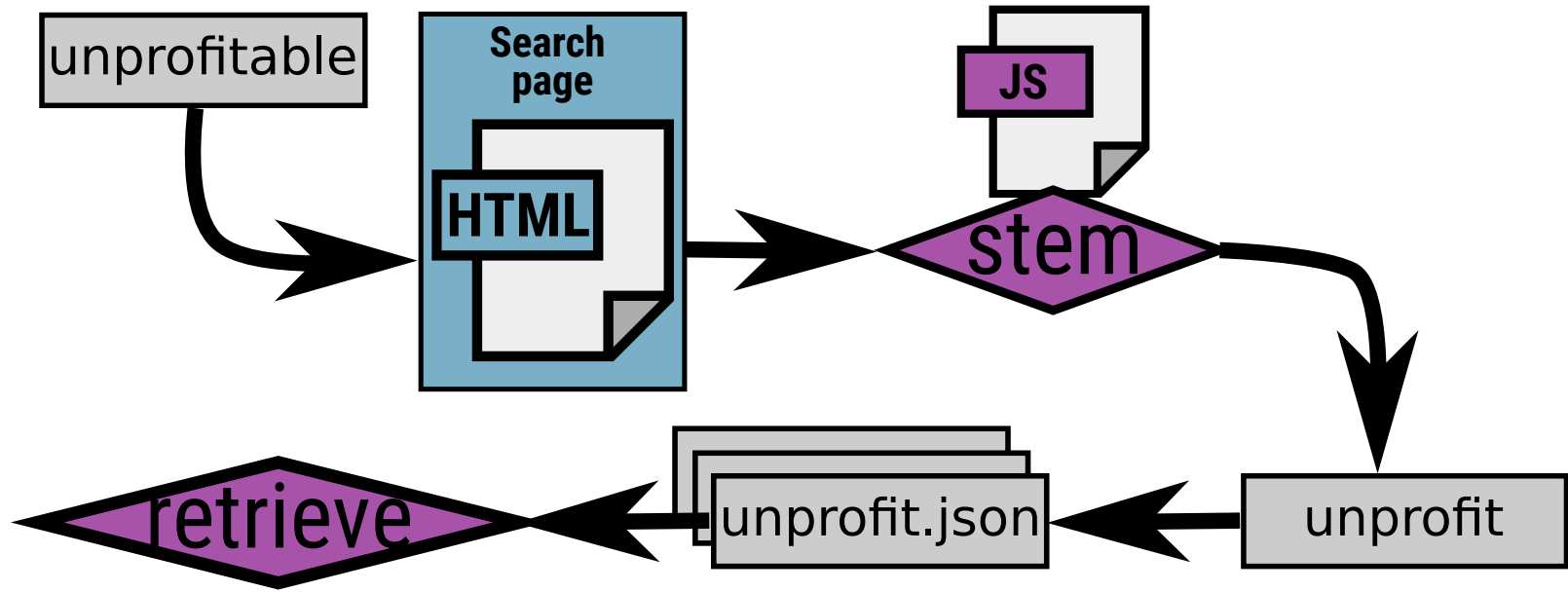


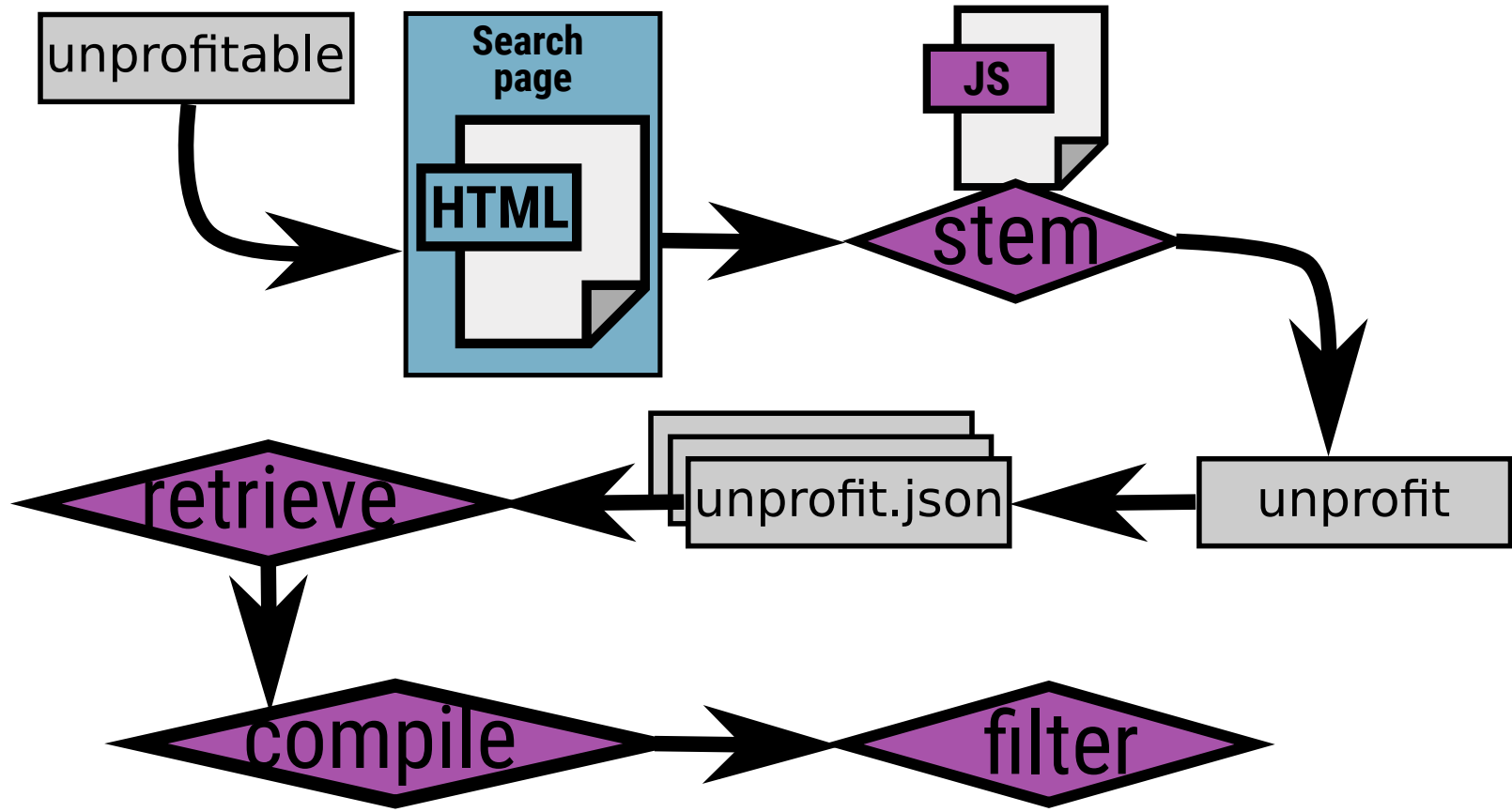
unprofitable

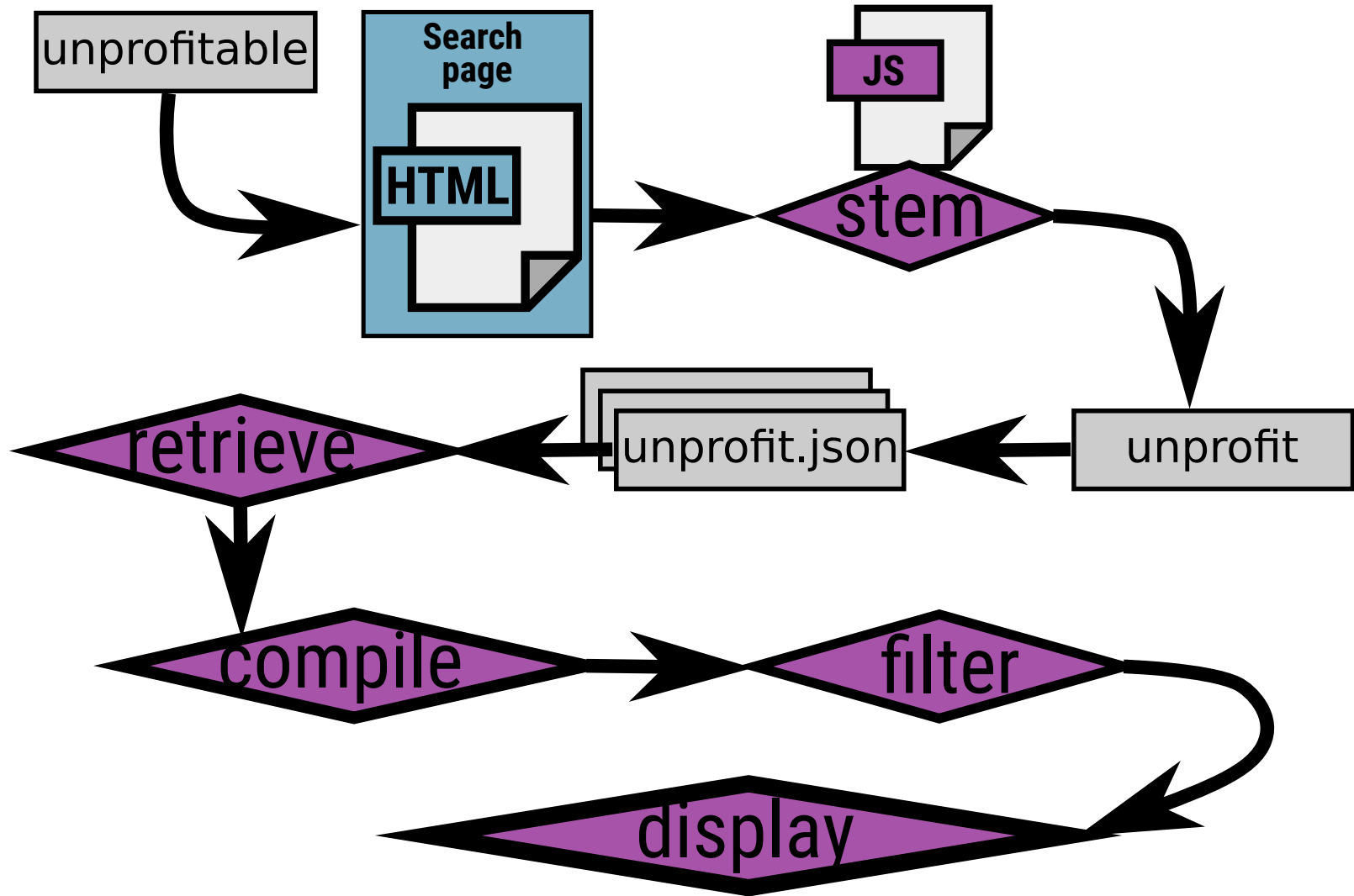












Documents found: 2

I.—The Old Man of Hoy Score: 1

- ...whole that day was spent unprofitably.

John and Joan, Canto II Score: 1

- ...too much ap- propriated unto unprofitable jocularities and facetiousness. Craving licence,...